

# Las pruebas estandarizadas de evaluación

## Un enfoque crítico y propositivo



María Teresa Sales | Licenciada en Ciencias de la Educación (UdelaR). Forma parte del comité académico de posgrado ANEP-UDELAR en Didáctica. Integrante del GRE (Grupo de Reflexión sobre Educación).

### Algunos puntos de partida

Ningún instrumento de evaluación en lo social/educativo es neutro. Son prácticas sociales construidas históricamente, y responden a procesos estructurales o coyunturales ligados a demandas que provienen con frecuencia de intereses y lógicas de otros campos, exógenos a una educación con centro en la persona y su sociedad de pertenencia. Por ese motivo es necesario analizar críticamente los dispositivos que surgen o se *aggiornan*. Este último es el caso de las pruebas estandarizadas. Consisten básicamente en las llamadas pruebas objetivas y, como todos sabemos, son diseñadas y se aplican en versiones nacionales e internacionales.

No abordaremos en este artículo el aspecto técnico de las pruebas estandarizadas, es decir, su metodología de diseño, su validación, su aplicación y el análisis estadístico de los datos recabados por estos dispositivos. Nuestros cuestionamientos a estas pruebas no radican en los mecanismos técnicos de su producción y posterior tratamiento, propios de la psicometría, que presuponemos que son metodológicamente correctos.

Lo que motiva este artículo son las razones de su auge actual en buena parte del mundo, su lógica e ideología de base, sus limitaciones en relación a lo que se plantean medir, así como sus potenciales riesgos y efectos constatables en aquellos países que vienen transitando por su aplicación con carácter certificativo. Nos planteamos, asimismo, qué papel podrían tener, integradas en una evaluación que tenga en cuenta la multidimensionalidad y complejidad de los aprendizajes.

De lo anterior surge la intención teórica y ético-política de este trabajo: cuestionar la neutralidad proclamada de estos instrumentos para contribuir a problematizarlos en lugar de “darlos por buenos”, así como para recuperar su eventual potencial en términos de una educación concebida como derecho humano y no como mercancía.

Todo instrumento de evaluación aplicado a estudiantes implica siempre una intervención sobre el evaluado, ya sea directa cuando son certificativos, o indirecta porque responden a la vez que promueven a una determinada corriente pedagógico-didáctica que parte de una determinada concepción antropológica y sociológica, incide en la formación de las personas y tiene, por lo tanto, consecuencias sobre sus vidas y sobre la sociedad.

En su versión internacional, algunas pruebas estandarizadas responden a necesidades que no son propias del campo educativo y repercuten en las naciones.

*«Inadvertidamente, cuando un gobierno acepta sumarse al Programa acepta que el criterio de medida de los logros de aprendizaje de sus alumnos será, no el currículo nacional, sino el marco de evaluación de competencias establecido por PISA. La asunción que entonces se realiza es muy simple: hay que mejorar la posición relativa del propio país en el concierto internacional porque de ello depende la competitividad económica. Es decir, el criterio de medida se convierte inevitablemente en un objetivo político per se, sin que haya mediado necesariamente ninguna discusión entre los actores a escala nacional (Grek, 2009). Es el advenimiento de la gobernanza global de la educación.»* (Pedró, 2012:153-154)

Como veremos más adelante, PISA y otras pruebas internacionales no actúan unilateralmente en su proyecto homogeneizador, sino que, para nuestro caso, establecen con OEI y sus Metas 2021 una acción cruzada en torno a la evaluación en relación a la enseñanza por competencias que impone la Meta General Quinta a los países latinoamericanos que las suscribieron, entre los cuales se cuenta Uruguay.



Desde la perspectiva de ciencia social crítica en la que este artículo se ubica, todo dispositivo responde en forma predominante a un modelo evaluatorio que se sustenta en supuestos ontológicos, antropológicos, sociológicos, epistemológicos, pedagógicos, éticos y político-económicos que de alguna manera lo “formatean”. Esos supuestos no son verdades fuera de tiempo y espacio, sino que son en sí mismos productos históricos. Su génesis, su cuestionamiento, su reaparición o predominio, con frecuencia cíclica a lo largo del tiempo, responden a razones tanto ligadas a cambios culturales profundos (de corte paradigmático) como a requerimientos provenientes de campos externos al educativo, entre los cuales el económico-productivo, con sus transformaciones, sus crisis y renacimientos, ha constituido, desde los orígenes de los sistemas educativos modernos en el siglo XIX, la gran fuente de demandas a la educación.

### 1. Vehiculizando “buenas intenciones”

Cuando el enfoque del campo de la evaluación obvia la consideración de los aspectos de fondo que subyacen en los instrumentos y de sus potenciales efectos sobre las personas, y se limita en cambio a los aspectos técnicos de las pruebas por sus promesas de objetividad y de eficiencia, ya que permiten abarcar a grandes grupos con menos recursos, de eficacia y validez para “medir” lo que se proponen, de confiabilidad por el carácter “científico” de esas mediciones estadísticas, o incluso por su valor “formativo”, se vehiculizan una serie de “buenas intenciones” de variada gama, que se naturalizan y que, por lo tanto, tampoco se discuten. Por su parte, el empleo actual de la tecnología informática les aporta un valor agregado, ya que promete facilitar el logro de aquellas buenas intenciones, pero en realidad contribuye a invisibilizar sus supuestos de partida, sus fines y sus funciones, que no son neutros en términos ideológicos.

Las pruebas estandarizadas suelen ser valoradas por su supuesta objetividad como “el” instrumento para diferentes fines: el *ranqueo* de países en el todo mundializado (función de las pruebas internacionales), la “*accountability*” (rendición de cuentas a la sociedad sobre la calidad de un sistema educativo), la *regulación* y, en algunos países, la *certificación* de los desempeños de los estudiantes.

**1.1.** En la globalización económica y cultural se pone el énfasis en la competitividad de los países en términos del “capital humano” de que disponen para el crecimiento económico. Los *rankings* internacionales establecidos a partir de las pruebas estandarizadas buscan precisamente clasificar a los países en relación a esa competitividad. La educación es entonces considerada insumo privilegiado para la productividad del campo económico, se mercantiliza y pasa a ser tratada como un bien de inversión y de cambio, y no un derecho humano y un bien público.

**1.2.** La “*accountability*” integra la lógica del llamado Estado Evaluador, figura que surge luego de la crisis económica de los ochenta y se solapa, con diferente peso según los países, con la lógica de dador y administrador burocrático, propia del Estado de Bienestar del período anterior a la crisis. Para algunos autores constituye un nuevo contrato social, en el que el Estado, en sus versiones más neoliberales, pasa a ser prioritariamente evaluador, en este caso, de los resultados de los aprendizajes del sistema educativo, los que son considerados la medida de la calidad del sistema. Este debe de rendir cuentas, en una extrapolación de la lógica financiera de costo-beneficio (este último lo más pronto posible), para la competitividad a que aludíamos en el numeral anterior.

Malos resultados habilitan directa o indirectamente a tomar decisiones que en algunos países con frecuencia son marginadoras de estudiantes, docentes e instituciones en una lógica propia de un darwinismo social basado en la libre competencia.<sup>1</sup>

**1.3.** El monitoreo o regulación de los resultados de los estudiantes mediante las llamadas pruebas estandarizadas o tests no ha tenido ni tiene una misma y única función en los diferentes sistemas educativos. En nuestro país, estas pruebas en la educación básica comenzaron en 1996, en tanto en 2005 pasaron a ser aplicadas en línea. Tuvieron en origen función “diagnóstica” y de aplicación voluntaria. Recientemente se agregaron pruebas “formativas”, vehiculizando otra buena intención: la de promover la reflexión en el cuerpo docente. Por definición teórica del modelo, esta reflexión se centra en los resultados de los estudiantes, para eventualmente realizar trabajo remedial y revisar el abordaje pedagógico-didáctico realizado en función de los objetivos prefijados. Se trata de la función de retroalimentación o *feed back* de la evaluación, dentro de la perspectiva tecnicista de la enseñanza.

Si bien en algunos países, los tests o pruebas estandarizadas se aplican desde hace más de cien años (en los EE. UU. concretamente), en algunos países como Canadá, los propios Estados Unidos, Inglaterra, Gales y la India entre otros, las pruebas estandarizadas han derivado en los últimos años en cumplir una función certificativa con importantes repercusiones, por lo cual se las suele calificar como de “alto impacto”<sup>2</sup>. Se pasa (o no) de año, se accede (o no) a otro nivel del Sistema Educativo Básico, se accede (o no) a la Universidad, a través de la sucesión de múltiples pruebas estandarizadas pautadas en el curso de la formación.

<sup>1</sup> Se puede acceder a mayor información sobre la evolución en las últimas décadas del rol del Estado en la regulación de los sistemas educativos en Bentancur (1997-1998), y en Mottier y Crahay (2009).

<sup>2</sup> Traducción libre de “*high-stakes tests*”.



Hoy, en algunos países insertos en el modelo neoliberal o de cuasi mercado educativo, el evaluar a los docentes y a las instituciones a través de los resultados de los estudiantes en estas pruebas es un insumo privilegiado para medir la eficiencia, productividad y eficacia de unos y otras, y para distribuir premios y castigos “equitativamente”, siempre en el plano de otra “buena intención”: mejorar la calidad de la educación. Obviamente, la necesidad de una educación de calidad no es discutible, pero la calidad no tiene un único sentido. Si consideramos que la calidad es el conjunto de características de una educación concebida como derecho humano y bien público, como la definía la consultora de UNESCO, Katarina Tomasevski, no se la puede limitar a ser prioritariamente insumo del campo económico y sus avatares, lo cual la reduce a una función: la económica.

En realidad, el empleo de la función certificativa de las pruebas de desempeño de los estudiantes y sus efectos sobre otras dimensiones educativas, responden a las razones de fondo que ya presentamos en los puntos 1.1. y 1.2. de este apartado. Su aplicación en los sistemas ha causado una serie de efectos colaterales nefastos, puestos en evidencia por prestigiosos autores como Figari y Mottier (2006); Hursh (2008); Nichols y Berliner (2010); Harris, Smith y Harris (2011); Giroux (2012). Volveremos sumariamente sobre este aspecto más adelante.

## 2. ¿Cuál es la lógica implícita de las pruebas estandarizadas?

Estas pruebas se ubican en el modelo cuantitativo de evaluación, para el cual “evaluar es sinónimo de medir”, y si la medición está bien hecha técnicamente, garantiza la objetividad de la evaluación. Esta es la perspectiva recogida por Luaces (2014:19) en el libro *Evaluaciones en línea*. Cabe precisar que esta concepción no es la única. Habría que investigar además si es la “comúnmente aceptada”, como se postula en el mismo libro. Incluso si así fuera, correspondería preguntarse si “debería ser” la comúnmente aceptada, a la luz de desarrollos teóricos actualizados y con sólidos fundamentos que cuestionan la asimilación del término evaluación educacional con el de medición. Esta asimilación se encuadra en el paradigma neopositivista largamente cuestionado por la filosofía contemporánea, la nueva sociología de la educación y la pedagogía, desde paradigmas alternativos de largo y fecundo desarrollo: el interpretativo y el crítico.

Tanto en las “pruebas en línea” nacionales como en las pruebas internacionales, entre las cuales se destaca PISA como instrumento de OCDE exportado mediante un marketing agresivo a una buena parte del mundo, se trata *prioritariamente* de pruebas de múltiple opción y/o de verdadero/falso. Es el resurgimiento generalizado, ya sin la apoyatura explícita del Conductismo hoy teóricamente superado, de las llamadas “pruebas objetivas”<sup>3</sup>. El calificativo “objetivo” presupone su valor: se asume que este tipo de pruebas, si están bien elaboradas y bien fijados estandarísticamente sus estándares, “miden” lo que un estudiante “sabe”, o “sabe hacer” (sus competencias) en un determinado nivel de un campo de conocimiento y pueden ubicarlo en el “nivel” que le corresponde.

La presunción de objetividad elimina la incertidumbre a través del número, validador de gran prestigio en las sociedades occidentales contemporáneas, en las que se observa una obsesión por las mediciones, las estadísticas y los *rankings*. En esa línea de pensamiento se puede caracterizar como “objetiva” la constatación o no de un aprendizaje cuando se basa en hechos observables y no es influida por opiniones, emociones o prejuicios del evaluador. Aprendizaje en este sentido es sinónimo de desempeño. Correlativamente, el modelo asume que es posible transferir la irrenunciable responsabilidad ética del juicio sobre el “otro” que consideramos tiene todo evaluador en el campo social, a una metodología “científica” de diseño de instrumentos de medida de los aprendizajes.

Más allá de la discusión epistemológica contemporánea en relación a la posibilidad misma de objetividad, que no abordamos en este artículo ya que pretendemos no ubicarnos en el plano de la reflexión abstracta, el cuestionamiento y la relativización de esta lógica objetivista concretamente en las pruebas estandarizadas pueden ser planteados por los menos en tres planos sustantivos: a qué responde la selección de los objetos a evaluar; qué “miden” y qué no miden las pruebas objetivas estandarizadas; y a qué fines sirven o pueden servir eventualmente, sus efectos y consecuencias.

<sup>3</sup> Se considera vigente al conductismo y a una didáctica acorde, específicamente en aquellos casos en los que se requiere automatizar y consolidar mediante mecanismos de estímulo-respuesta, comportamientos adecuados frente a requerimientos profesionales instrumentales, o por razones de seguridad industrial, propios por lo general de la capacitación técnica.



## 2.1. Los objetos de evaluación

Las pruebas estandarizadas, ya sean nacionales o internacionales, responden a una selección del objeto a evaluar que no es objetiva en sí misma. El currículo escolar de un país, en su sentido amplio, que comprende lo que se prioriza a través de las evaluaciones y forma parte del currículo oculto, es el primer recorte dentro de todo el conocimiento, las habilidades, las competencias y los valores existentes en un momento y tiempo dados.

Si bien toda propuesta evaluatoria, incluso la elaborada por cada docente de aula, implica un segundo recorte, tampoco objetivo de conocimientos y de competencias, las pruebas estandarizadas al generalizar los mismos objetos a evaluar, establecen ese recorte preceptivamente para toda la población escolarizada.

Las limitaciones de la objetividad son mayores aún en las pruebas internacionales, en las que a pesar de algunos esfuerzos de contextualización, ligados a reclamos de países subdesarrollados, son comparados países de culturas, tradiciones educativas y currículos diferentes, y se los ordena sin embargo en un *ranking* mundial.

Hay una velada imposición curricular en estas pruebas, que ya opera en algunos países. La forma en que los estudios PISA miden los logros de los alumnos tiene por lo menos la vocación de convertirse, de manera sutil y progresiva, en un prescriptor de los objetivos de las políticas educativas nacionales, al establecer los estándares de competencias de los alumnos de quince años de edad para todos los países desarrollados que participan y para los que aspiran a desarrollarse. Parece entonces de orden preguntarse por qué tanto PISA como las pruebas nacionales en línea, siguiendo a PISA, se proponen medir los resultados de competencias solo en matemática, ciencias (naturales) y lengua para evaluar al sistema. ¿Por qué no letras, ciencias sociales o arte en su lugar, o por lo menos, además de aquellos tres campos?

Si bien se suele fundamentar la selección en el carácter de instrumento de las tres áreas para otros aprendizajes (otra “buena intención”), también es cierto que por lo menos matemática y ciencias forman parte del conocimiento de alto estatus, por ser básicos en la productividad y el crecimiento económico.

La objetividad de las pruebas estandarizadas está entonces sesgada desde el origen, ya que responden en los hechos a medir respuestas a imperativos económicos, hoy mundializados, lo que constituye un arbitrario cultural, en términos de Bourdieu.

También los ítems mismos de las pruebas están inevitablemente teñidos de la subjetividad de quienes, con formación y la mejor de las intenciones, los seleccionan y elaboran, lo que no necesariamente coincide con el currículo “vívido” por los docentes que trabajan diariamente con los estudiantes, en el que hacen otra selección de contenidos aun del mismo currículo nacional prescrito, otras prioridades, otras formas de enseñar y de evaluar durante el curso. Paradojalmente, el desiderátum de la objetividad a

través de la medición lleva nuevamente, como sucedió en el conductismo más ortodoxo, a considerar al aula como una caja negra inexplorable, por lo cual las mediciones científicas no pueden considerar lo que pasa en ella. En la medición de resultados de las pruebas estandarizadas nacionales, por definición teórico-metodológica del modelo cuantitativo de evaluación que las sustenta, no es posible tener en cuenta lo que sucedió en el proceso pedagógico didáctico, que es nada más ni nada menos donde se construyen conocimientos y competencias. Se trata pues de una medición por fuerza imperfecta. Esto, que es una limitación en las pruebas nacionales no certificativas, repercute sin embargo en un ordenamiento de las instituciones que valoramos como muy positivo que se haya resuelto que no se haga público, pero hemos comprobado personalmente que lamentablemente “se filtra”. En cuanto a los desempeños de los estudiantes ha sido investigada la incidencia, en los resultados de estas pruebas, del hábito o no de responder a pruebas estandarizadas, así como la capacidad para adaptarse a los cambios y al estrés que las mismas puedan implicar. Son aquellos estudiantes que en sus vidas cotidianas son estimulados para enfrentar nuevas situaciones, los que mejor se desempeñan en este tipo de pruebas.

*«Una de las razones principales por las cuales el status socioeconómico de los niños está tan altamente correlacionado con los puntajes de las pruebas estandarizadas es que muchos ítems de las pruebas estandarizadas de logros se concentran en realidad en evaluar conocimientos y/o destrezas aprendidas fuera del colegio – conocimientos y destrezas probablemente aprendidas más en algunos escenarios socioeconómicos que en otros.»* (Popham, 1999:9)

A nivel internacional son los países con más larga tradición en pruebas externas los que responden mejor a PISA, e incluso incide en los resultados la tendencia de los estudiantes de algunos países a no detenerse en los ítems difíciles y seguir avanzando hasta el final, con lo que aventajan en *cantidad* de respuestas a los estudiantes que como los españoles tratan de responder avanzando ítem tras ítem, demoran más y dejan ítems sin responder.

Los anteriores son solo ejemplos de la gran complejidad que caracteriza a la educación en contexto, por lo cual las diferencias que se constatan entre los resultados de estudiantes y de países no responden exclusivamente a los factores cuantificables que por reduccionismo teórico-metodológico toman en cuenta las pruebas estandarizadas nacionales e internacionales.



## 2.2. ¿Qué “miden” y qué “no miden” las pruebas objetivas estandarizadas?

El modelo subyacente de evaluación en la base de este tipo de instrumento presupone que solo lo que se puede observar se puede medir. Agregamos: y si no fuera así, razonando en forma simple, siempre quedaría el recurso de no evaluar aquellos aprendizajes que no son observables. Precisamente es lo que reconoce el libro *Evaluaciones en línea*: (La concepción de evaluar como sinónimo de medir) «*presenta el interés de que los datos pueden ser tratados matemáticamente y presenta la desventaja, que deja de lado en la evaluación todo aquello que no es posible ser medido*» (Luaces, 2014:19). Nuevamente se trata de un reduccionismo por el método que implica una selección de objetos a evaluar que, por eso mismo, no es objetiva ni neutra.

El problema de fondo a este respecto se plantea cuando entramos a pensar el valor de los conocimientos, saberes y competencias que se pueden medir y de los que no. Por supuesto que con una prueba de múltiple opción se puede medir el reconocimiento de contenidos curriculares transmitidos, el manejo de algoritmos, la aplicación de conceptos, la capacidad de lectura literal e incluso inferencial de un texto<sup>4</sup>. Y es válido hacerlo a título de información de base. No es tan clara la posibilidad de evaluar saberes, competencias (complejas y no meramente instrumentales) y actitudes cuya importancia en la formación de las personas es clara con su sola mención y que hacen a un pensamiento autónomo y a una cabeza «*bien hecha*» (Morin, 2007) a diferencia de una bien llena: analizar, sintetizar, fundamentar, contextualizar, relacionar, cuestionar, argumentar, integrar conocimientos, el pensamiento crítico, la creatividad, la curiosidad, la apreciación estética, incluso pensar matemáticamente, científicamente, analizar críticamente mensajes, pensar por problemas, como proponía Vaz Ferreira, y no solo darles solución, entre otras muchas.

Como al decir de Gimeno Sacristán, la evaluación condiciona todo, lo epifenoménico que las pruebas estandarizadas miden, orienta la reflexión posterior de los docentes al mismo nivel de análisis mediante el tratamiento de los errores que estadísticamente priman y sobre los cuales se brindan centralizadamente algunas de las posibles causas. Por un lado, los aprendizajes son procesos sumamente complejos, en los cuales los posibles errores son difícilmente reductibles a unas pocas causas. Por otro lado, con el análisis, de alguna manera estandarizado y en abstracto de los errores de un grupo, se olvida que son parte natural de procesos individuales de aprendizaje diferentes, ligados a tiempos, circunstancias e historias personales y a la maduración biológica; a la vez que se presupone, además, que las redes neuronales se producen de la misma forma en todos los individuos.

<sup>4</sup> Estas pruebas miden prioritariamente el identificar un conocimiento entre las cuatro opciones que brinda el ítem, lo que no es un procedimiento habitual en la vida corriente. Por otro lado no miden en forma prioritaria ni la evocación ni la reproducción personal de un conocimiento, las que implican procesos mentales más complejos y no se avienen al formato de los ítems de múltiple opción. Cabría investigar si los ítems de producción incorporados en las pruebas en línea lo hacen.

Deducir y promover linealmente cambios en la didáctica a partir del porcentaje de errores de un grupo en un ítem, abona las perspectivas de una supuesta homogeneidad de aprendizaje de los estudiantes y de una relación causal unidireccional entre la enseñanza y éstos últimos. Se trataría, por supuesto, de analizar las prácticas de aula, pero además se debería investigar cómo aprenden los estudiantes en situación e individualizar sus procesos, más que tratar prioritariamente de controlar y remediar los errores que aparecen estadísticamente en las pruebas. El problema es que estas últimas miden desempeños; y esto nos lleva al párrafo siguiente.

Las pruebas estandarizadas *tradicionales* no pueden contemplar los *procesos* de aprendizaje.<sup>5</sup> El calificativo de “evaluación formativa” fue acuñado por Scriven (1967) como alternativa a la tradicional y predominante evaluación sumativa, por considerarla insuficiente. No sustituye pero supera la concepción técnico-instrumental de evaluación ligada exclusivamente al *control* del logro de objetivos (Tyler, 1949) para proponer la mejora como fin evaluatorio. Si bien Scriven la refirió a la evaluación de programas educativos, con el tiempo se amplió el concepto a una evaluación de los aprendizajes que tome en cuenta no solo resultados, sino *procesos*, y promueva la reflexión de los docentes y de los propios estudiantes sobre avances, retrocesos y necesidades puestas en evidencia, que redunde en la construcción individual y colectiva de una enseñanza y de aprendizajes más personalizados y eficaces. Obviamente no se trata de una evaluación formativa cuando se incorpora una prueba a mitad de año, ya que se evalúan desempeños terminales, en este caso de un período, como sucede en la propuesta actual de ANEP. Lo que pasó en ese período forma parte de un proceso tomado como caja negra sobre el que se promueve la reflexión *a posteriori*, infiriendo solo los docentes y supervisores lo que sucedió (cf. Luaces, 2014:127 y ss.).

Para evaluar integralmente es necesario superar la lógica objetividad = medición, sin negar la posibilidad y pertinencia de la medida según el objeto a evaluar, pero introduciendo en un plano de igualdad lo cualitativo para la *comprensión* de procesos y de resultados, que son en sí mismos complejos y multicausales. Para ello es necesario potenciar las evaluaciones de los docentes e incorporar dispositivos abiertos de evaluación (los portfolios, la evaluación de competencias mediante situaciones complejas, por ejemplo), que no dejan de lado referentes, criterios y evidencias o indicadores, pero que permiten a docentes *con buena formación en evaluación* saber qué observar, cómo llevar adelante las observaciones y llegar a conclusiones para (re)orientar los procesos de aprendizaje.<sup>6</sup> Evaluar no se agota en ubicar las lagunas y dificultades de un alumno,

<sup>5</sup> Veremos en este mismo apartado que existen intentos en el mundo de transformar las pruebas tradicionales (o de primera y segunda generación), asignándoles otros sentidos y funcionalidades que integran la evaluación de procesos.

<sup>6</sup> Hemos trabajado una propuesta metodológica en esta dirección en Sales, Rodríguez y Sarni (2014:Cap. 4).



sino también y sobre todo implica comprender cómo estructuró los conocimientos e incluso visualizar el proceso de aprendizaje con el fin de identificar el o los momentos en los cuales los desfases se produjeron.

Ante la insuficiencia de las mediciones tradicionales estandarizadas de resultados, existen en otros países o están siendo ensayadas nuevas generaciones de pruebas en línea, que se orientan a evaluar y a reorientar procesos, esta vez individualizados, pasando de una evaluación sumativa a una formativa y diagnóstica para mejorar los aprendizajes, con el apoyo de cada vez más sofisticadas herramientas informáticas.

Nuestro país viene aplicando desde 2007, pruebas en línea de formato único, que se insertan en la denominada primera generación de evaluación asistida por computadora. Desde 2009 se trabaja en el diseño de los llamados Tests Adaptativos Informatizados (pruebas TAI) (Luaces, 2014:Cap. I) que constituyen una innovación (se ubican en la segunda generación de pruebas), ya que permiten determinar el nivel de desempeño de un sujeto en forma individualizada y con menos ítems, pero no permiten ni diagnosticar las razones de un logro o de un fracaso ni precisar los mecanismos puestos en juego (cf. Burton y Martin, 2006:155)<sup>7</sup>.

Las dos primeras generaciones de evaluación asistida por computadora sirven para una aproximación tanto sistémica como individual aunque, como vimos, con limitaciones, pero siempre son del tipo sumativo o incluso certificativo. Son empleadas con frecuencia a los efectos de ubicar estudiantes en distintos niveles, en relación a estándares de competencia predefinidos. Desde los años noventa se vienen empleando con ese fin clasificatorio tanto en los EE. UU. como, en menor medida, en Europa (Países Bajos, Luxemburgo), y se ha demostrado que con frecuencia esas clasificaciones han sido erróneas.

De acuerdo a Burton y Martin en el artículo de referencia, la tercera generación de tests informatizados es la denominada “de medida continua”, con un funcionamiento calcado de las pruebas TAI, “pero pensada para ser aplicada en continuo, de modo de ser integrada en el proceso mismo de aprendizaje con vistas a guiar a los estudiantes. La medida se vuelve dinámica dentro de la óptica de una evaluación formativa”. En la cuarta generación de pruebas, llamadas “de medida continua e inteligente”, se trata de que el sistema informático actualice permanentemente un modelo del estado cognitivo de cada estudiante, que permita brindar informaciones precisas para visualizar aquellas intervenciones pedagógicas que lo ayuden a progresar en sus procesos de aprendizaje. Se ha asimilado este abordaje de las pruebas a una especie de tutoría que, una vez efectuado el diagnóstico, pone a disposición del estudiante una variedad de recursos de información en formato multimedia sobre los problemas detectados. Una tercera etapa consiste en que el estudiante en un proceso de autoformación utilice el material a su disposición con el fin de retrabajar y

mejorar sus respuestas iniciales, todo lo cual es registrado en una plataforma que permite realizar análisis tipológicos de estrategias de exploración y empleo de la información. Se argumenta que si bien esta evaluación permitiría trazar el proceso de aprendizaje individual en un *continuum*, no sería capaz de diagnosticar en forma precisa el funcionamiento cognitivo de una persona en un momento dado y ligarlo con fundamento teórico a intervenciones pedagógicas adecuadas. Se apuesta a futuro a que la psicología cognitiva brinde la posibilidad de crear una “psicometría cognitiva” (*ibid.*, p. 161), lo que está recién en etapa de investigación.

Nos hemos detenido en el pasado, presente y el potencial futuro de las pruebas asistidas por computadora en países que las vienen aplicando desde tiempo atrás. En un intento de aproximación de lo “cuanti” y lo “cuali” en aquello que pueden ser epistemológicamente reconciliables, hemos acudido a autores que las defienden desde su lógica interna de medida, pero lo hacen con sólidos fundamentos y no meramente apuntando a la eficiencia y eficacia como el ahorro de recursos materiales y/o de tiempo. No son autores apologeticos: piensan en forma compleja y ponen de relieve, desde la investigación, las inconsistencias y los riesgos ya constatados de su aplicación acrítica.

Consideramos que explicitar, aunque brevemente, el recorrido que están haciendo otros países en relación a los tests es útil en varios aspectos: muestra su complejidad tanto teórica como instrumental, los cuestionamientos y la superación de su estandarización tradicional, de la fantasía de su potencialidad para medir aspectos fundantes (competencias implícitas, competencias sociales, resolución de problemas complejos en el marco de una ecología de saberes, etc.), la necesidad del empleo preciso de la calificación de “formativa” para una comprensión/medición necesariamente de procesos y no solo de resultados parciales.

De todas maneras, mucho más allá de lo instrumental, se requiere hacer foco en lo que realmente importa: la persona en etapa de formación y su sociedad de pertenencia, por qué y para qué evaluar, y en ese marco contrastar con los fines a los que responden hoy las evaluaciones.

### 2.3. A qué fines sirven las pruebas estandarizadas (o pueden llegar a servir)

La ciencia y especialmente sus aplicaciones tecnológicas no son neutras. Se encuentran vinculadas al sistema del poder y sus herramientas ideológicas y financieras. La globalización económica de la sociedad capitalista implica la competencia por el conocimiento y sus aplicaciones tecnológicas como herramienta de poder, en términos de la competitividad relativa entre países y bloques, para el crecimiento económico.

<sup>7</sup> Traducción y aclaración entre paréntesis de la autora.

Las pruebas estandarizadas internacionales tienen como objetivo establecer *rankings* entre países a partir del promedio de los resultados de sus respectivos estudiantes. En ese sentido, PISA es el instrumento con el que la OCDE “se propone incidir en las políticas educativas de su zona de influencia, hacerlas converger más, aumentar el desempeño y la equidad de los sistemas educativos en nombre de la competitividad, en una sociedad del conocimiento y un mundo globalizado” (Lafontaine, Soussi y Nidegger, 2009:78).

La pertinencia de las comparaciones internacionales es cuestionada por autores críticos tales como Harris, Smith y Harris, y G. Bracey, tanto por una multiplicidad de aspectos técnicos, como de sentido común. Harris, Smith y Harris (2011:18), razonando dentro de la lógica mercantil, sostienen que “...es claro que ninguna economía nacional es construida enteramente por ingenieros y científicos, por lo cual todo el emprendimiento de juzgar una economía por el promedio de los niveles logrados en los tests en matemática o ciencia o incluso en lectura es realmente un ejercicio infructuoso”. Bracey (cf. Harris, Smith y Harris, 2011) por su parte, argumenta que los efectos amplios y de largo alcance en educación no pueden ser reducidos a unos pocos y triviales indicadores y todo sistema educativo puede ser comprendido solo teniendo en cuenta su historia, sus fines, y la complejidad de sus estructuras.<sup>8</sup>

El problema mayor en los diferentes países se da cuando esos resultados, tomados como representativos de la realidad, se traducen en *rankings* de instituciones en nombre de la “buena intención” de transparencia para la rendición de cuentas a la sociedad, en relación a una calidad educativa que en forma circular se define como aquella que contribuye a la productividad, la libre competencia y el crecimiento económico como bienes en sí mismos interrelacionados. De esa manera, la lógica del modelo busca autosustentarse.

En los estudiantes, el problema mayor se plantea cuando estas pruebas dejan de ser indicativas o diagnósticas y pasan a ser certificativas; es decir, cuando se califica y se certifican los cursos o incluso los niveles del sistema educativo a través de los resultados en ellas, repercutiendo directamente en la vida de los estudiantes. A partir de esos resultados se diseñan políticas educativas, o incluso sirven, como ya adelantáramos, para la evaluación de los docentes y de las instituciones educativas, lo que es acompañado de premios y de castigos que pueden llegar a la no renovación del contrato del docente, la supresión de subvenciones institucionales e incluso el cierre de algunas escuelas por descenso de la matrícula, dado que son las familias quienes eligen la institución, en un régimen de libre competencia. Estas consecuencias de las pruebas estandarizadas a nivel macro son hoy el instrumento privilegiado de regulación de la educación en varios países, en un régimen entendido como de cuasi-mercado, prolongación actual de las políticas neoliberales de los noventa.

<sup>8</sup> Traducciones de la autora.

Si bien nuestro país no ha entrado en esa vía, la presión mundial hoy existente en relación a la rendición de cuentas con fines de *ranqueo* y clasificación de los países en la competencia internacionalizada, las tendencias homogeneizantes que se generan mundialmente con ese mismo fin y la interrelación de ambos procesos entre sí, está en la base de formas indirectas de incidencia de este tipo de pruebas en la educación de los países.

Esa incidencia viene por la vía ya no solo de exigencias de los tradicionales organismos internacionales de crédito con injerencia en la educación (BM, FMI), sino también de organizaciones supranacionales de países, que se retroalimentan entre sí: la OCDE europea y la OEI. Esta última organización legitima las pruebas nacionales y las internacionales al remitir a ellas la evaluación de las competencias, consideradas como indicadores de la calidad educativa de un país<sup>9</sup>, lo que lesiona la autonomía de la educación nacional consagrada constitucionalmente.<sup>10</sup>

Sería indispensable que el país analizara críticamente su participación en este tipo de proyectos, a la luz de fines y objetivos propios consensuados. Dentro de las Metas 2021<sup>11</sup> de OEI, en la Meta General Quinta: “Mejorar la *calidad* de la educación y el currículo escolar” se establece la siguiente Meta Específica No. 10: *Mejorar el nivel de adquisición de las competencias básicas y de los conocimientos fundamentales por parte de los alumnos*, cuyo indicador es el porcentaje de alumnos con niveles satisfactorios de logro en competencias básicas *en las pruebas nacionales e internacionales*.

Se fija el nivel de logro del indicador anterior en que: *Disminuye en al menos un 20% el número de alumnos situados entre los dos niveles bajos de rendimiento en las pruebas de LLECE 6.º grado, PISA, TIMMS o PIRLS en las que participan diferentes países. Aumentan en la misma proporción los alumnos en los dos niveles altos en dichas pruebas*.<sup>12</sup>

<sup>9</sup> La incapacidad de evaluar la calidad educativa a través de las pruebas que nos ocupan es desarrollada en Popham (1999).

<sup>10</sup> Se encuentra un análisis global y detallado de las Metas 2021 de OEI (Organización de Estados Iberoamericanos) en sí mismas y en correlación con políticas sustentadas por el Banco Mundial, en el artículo de Pablo Imen “Un Proyecto Educativo Neocolonial para Nuestra América. Contribución a la crítica de las Metas 2021 de la Organización de Estados Iberoamericanos” en Revista Voces, N° 7. Montevideo, julio de 2013.

<sup>11</sup> En línea: <http://www.oei.es/metas2021.pdf> (p. 152)

<sup>12</sup> La meta de OEI citada coincide con el comienzo del trabajo por competencias en Primaria, planteado por la circular N° 22 de marzo de 2014 de la Inspección Técnica del CEIP, que formó parte del temario de la ATD de Piriápolis de junio 2014. Paralelamente, el país se ha integrado a la Red Mundial de Aprendizajes que patrocina el Banco Mundial, a través de la cual se vienen difundiendo materiales para la evaluación de competencias a través de rúbricas con estándares prefijados y homogeneizantes. También en Secundaria, algunas de las Inspecciones de Asignatura están promoviendo el trabajo por competencias. Algunos desarrollos primarios sobre el tema, que exceden la finalidad de este artículo, pero que lo complementan, figuran en SALES, M. T., “Informe a la ATD de Primaria”, 14 de junio de 2014, Piriápolis (distribuido oportunamente por la respectiva Mesa Permanente).



### 3. Pruebas estandarizadas: algunas potenciales utilidades, efectos y riesgos asociados para la educación

A este respecto cabe distinguir las pruebas nacionales de las internacionales. Estas últimas apuntan a medir la calidad de los sistemas educativos participantes y a establecer un *ranking* comparativo entre países. Las críticas a su potencialidad para cumplir esta función son bien conocidas: se trata de pruebas que no contemplan, porque no pueden hacerlo, los contextos culturales, sociales, económicos y educativos de cada país.

El problema es que PISA y, por lo general, las pruebas estandarizadas internacionales son, como argumenta Popham (1999:5), “de talla única”, pero más allá de las diferencias contextuales de tipo social y cultural, hay también diferencias en los fines y objetivos curriculares en cuanto a lo que se prioriza en las prácticas y en cómo se enseña en cada país. Los resultados nacionales en esas pruebas son por lo tanto altamente discutibles, a pesar de su solidez técnica.

Las pruebas estandarizadas nacionales, a pesar de las limitaciones de base ligadas a su lógica, algunas de las cuales son reseñadas en este artículo, pueden sin embargo ser útiles, al permitir una aproximación gruesa a los resultados de cada alumno y grupo en un campo de conocimiento determinado, en relación a la media del país. Potencialmente, esta aproximación estadística podría llevar al colectivo docente institucional a investigar en su escuela. Se podría así superar el nivel de las percepciones u opiniones personales/grupales, que suelen surgir en la reflexión post prueba, implementando procesos con anclaje institucional a los que el análisis de las prácticas y la investigación pueden aportar mucho. Se requiere, por supuesto, contar con docentes formados para tales fines. La posibilidad de reiterar la aplicación de una misma prueba por parte de los docentes puede ser útil para aportar en dichos procesos, así como para el monitoreo de alumnos con dificultades. Asimismo, la repetición de una prueba, pasado un tiempo de la primera, puede brindar insumos a los docentes y a los propios alumnos de los posibles cambios operados, promoviendo la reflexión conjunta de ambas partes.

Los efectos negativos y los riesgos mayores de este tipo de pruebas, tanto nacionales como internacionales, se dan cuando a las mismas se les asigna un “alto impacto”, es decir, cuando como ya vimos, a partir de ellas en forma unilateral se toman decisiones que afectan a los alumnos (cuando las pruebas se vuelven certificativas) y también, en cadena, a docentes e instituciones. Solo a título informativo y de prevención, dado que hasta el momento nuestro país no transita por ese camino, mencionamos algunos de sus principales efectos y riesgos asociados, ya constatados por la investigación en países que aplican estas pruebas de alto impacto en función de la rendición de cuentas.



- ▶ Los estudiantes perciben las pruebas como el fin último de su educación. Los docentes, por su parte, dado que su estatus profesional depende de los resultados de los estudiantes trabajan en función de elevar los mismos, con lo cual la educación se distorsiona.
- ▶ Se adiestra a los estudiantes para responder a este tipo de prueba, con lo cual la enseñanza y el aprendizaje profundo de los temas quedan minimizados. Los estándares altos, el nivel de exigencia y el impacto de estas pruebas en la progresión académica de los estudiantes provocan efectos marginadores. Solo los estudiantes que pueden afrontar los costos acuden a tutores particulares, con lo cual el sistema reproduce las diferencias sociales y culturales.
- ▶ Se estrecha el currículo prescrito, orientándose los docentes a enseñar los contenidos que se intuye o se sabe por experiencia de tests anteriores que van a ser objeto de medición.
- ▶ Salvo los casos en los que *a posteriori* se emplean estas pruebas para re-pensar la enseñanza, los tests constituyen una etapa cerrada, hasta sin devolución a los estudiantes.
- ▶ “Cuanto más se use cualquier indicador cuantitativo para la toma de decisiones en lo social, más va a ser objeto de presiones hacia la corrupción” (Campbell, 1975 *cf.* Nichols y Berliner, 2007:26). A partir de la profundización y generalización de las pruebas de alto impacto en los EE. UU., el fraude en el sistema educativo se da en múltiples actores, dimensiones y niveles del mismo antes, durante y después de las pruebas. Tiene diversas finalidades: presentar mejores resultados de los estudiantes, a los efectos, en el caso de los docentes, de evitar problemas en su carrera profesional, lograr un premio en forma de bonus por productividad, o incluso evitar que un estudiante con problemas pierda una prueba.
- ▶ La presión que se ejerce por medio de estas pruebas constituye el eje de una teoría de la acción: se presume que motivan, incentivan y logran mejores resultados, aunque la investigación ha mostrado que a largo plazo no es así. El fraude suele ser también una forma de resistencia por parte de estudiantes, docentes y administradores críticos. Finalmente, el Estado también engaña.<sup>13</sup>
- ▶ Las consecuencias de la aplicación de las pruebas estandarizadas de alto impacto constituyen para algunos autores, la “real crisis en la educación norteamericana” (Hursh, 2008 [subtítulo del libro]).

<sup>13</sup> El tema de diferentes niveles y alcances de los fraudes fue investigado y es desarrollado en profundidad y minuciosamente por Nichols y Berliner (2007:Cap. II).

#### 4. «Generar resistencias a lo que viene, supone tener alternativas»<sup>14</sup>

Una precisión en relación al título de este apartado. Cuando hablamos de generar resistencias nos referimos a lo que potencialmente puede venir impuesto en un futuro próximo, de la mano de corrientes mundiales que ya hemos mencionado como riesgos. En ellas se apuesta a las pruebas estandarizadas como forma de certificar resultados de aprendizajes con fines de control y *ranqueo* de personas instituciones y países, se descalifica la profesión docente y se deteriora su autoimagen sustituyendo sus formas de evaluar, por lo general insertas en los procesos didácticos, en función de una supuesta objetividad garantizada por su externalidad. Hasta ahora nada de esto está planteado así en nuestro país, y es necesario precisarlo. Sin embargo es imprescindible diseñar *participativamente* una política evaluatoria propia, que no adopte acríticamente “modas” que nos vienen dadas, a veces impuestas por acuerdos supranacionales cuyos fines son por lo menos discutibles.

El campo de la evaluación es particularmente sensible a esas corrientes, sobre todo en época de real o de hiperdimensionada crisis educativa, ya que ha funcionado tradicionalmente como mecanismo de control y ajuste de los sistemas educativos.

La evaluación es un campo de tensiones entre binomios con frecuencia considerados antinómicos por sus fines, sus funciones, sus lógicas, sus actores, los respectivos objetos de evaluación: evaluaciones para el control o para la mejora; evaluaciones externas-evaluaciones a cargo de los docentes de aula; evaluaciones internacionales-nacionales; modelos cuantitativos-cualitativos de evaluación; evaluaciones para el seguimiento de los aprendizajes-para el monitoreo y control del sistema; evaluaciones formativas-sumativas; evaluaciones de contenidos cognitivos-evaluaciones de competencias...

En este apartado intentamos perfilar algunas alternativas que, desde una perspectiva de la Complejidad, buscan articular algunas de las disyunciones anteriores, sin que ello signifique una postura epistemológicamente ecléctica de nuestra parte.

En relación a lo anterior, nuestra búsqueda de alternativas integra a las pruebas estandarizadas *nacionales* o *certificativas* por su *potencial*, aunque limitado, para constituir *uno* de los posibles insumos para evaluar (no para calificar) a los estudiantes con fines de acompañamiento, así como para promover, bajo ciertas condiciones institucionales, reflexionar e investigar *con rigor* sobre los procesos de enseñanza y de aprendizaje contextualizadamente en las instituciones, y para indicar eventualmente la existencia de algunos problemas en el aula y el sistema.

Consideramos, en cambio, que por los problemas de sesgo ideológico, su descontextualización y su vocación homogeneizante para la formación en competencias instrumentales del “capital humano” competitivo necesario en la globalización, no es pertinente la participación del país en pruebas internacionales y, especialmente, en PISA. A eso responde que el GRE (Grupo de Reflexión sobre Educación) al que pertenezco haya dicho NO en forma categórica y urgente a la participación de Uruguay en el programa PISA y en su versión en preparación para países en desarrollo (*PISA for Development*) (GRE, 2014:11 [num. 44]).

A continuación, algunos de los supuestos que consideramos deben estar en la base de una política evaluatoria.

#### 4.1. De la rendición de cuentas a las responsabilidades<sup>15</sup>

Corresponde cuestionar a la rendición de cuentas, hoy internacionalizada, como fin de la evaluación de los sistemas educativos, a partir del presupuesto ideológico de que las instituciones y los docentes no cumplen su papel social y cultural, y son “los” responsables de la “crisis educativa”. Al respecto suscribimos las opiniones de esta autora en el sentido de que:

*«No es que las escuelas no se han hecho responsables antes de que la ola sobre la rendición de cuentas se desarrollara, sino que se comprometen con los problemas que consideran importantes (algunos, es cierto, muy difusos) y se organizan a partir de ellos en mayor o en menor medida para diseñar una acción colectiva. Aquí las condiciones organizacionales y laborales constituyen un aspecto clave. En segundo lugar, una escuela no puede ser responsable hacia el exterior (sean estos los padres, la comunidad, el contexto) si primero no es responsable “hacia adentro” por sus procesos y resultados (Elmore, 2003). Esto implica abordar los efectos de la escolarización de niños y jóvenes a partir de un proyecto compartido por un colectivo docente. Y no hacemos referencia aquí a un proyecto educativo (o como se denomine) formalizado solo en un papel para entregar a la autoridad superior, sino a la propuesta que se concreta en el mediano plazo, y que se construye en las acciones cotidianas en el corto plazo, para asegurar que los alumnos o estudiantes aprendan contenidos social y subjetivamente relevantes. Esto supone acuerdos internos sobre objetivos, expectativas y prácticas orientadas a la mejora académica, es decir una propuesta coherente y pertinente. Además requiere capacidad de orientación y liderazgo de equipos directivos y estructuras organizativas que permitan concretarla, tanto propias de la institución escolar como externas a ella (apoyo sostenido y orientado a trabajar en torno a los problemas concretos que surgen en la escolarización de niños y jóvenes).» (Poggi, 2008:3-4)*

<sup>14</sup> Anderson (2012).

<sup>15</sup> Poggi (2008:3-4).



El orientarse por los conceptos de responsabilidad hacia adentro de la escuela y hacia su comunidad, para lo cual la dinamización de los Consejos de Participación sería de capital importancia, llevaría a una evaluación contextualizada que busque comprender los problemas institucionales y pedagógico-didácticos en situación para, a través de colectivos docentes reflexivos, intentar superarlos. Lo anterior ya deja planteada la necesidad de una sólida formación docente inicial y continua, a la que solo se puede llegar asignándole el carácter universitario que integra las funciones de docencia, investigación y extensión, y puede contribuir a la resignificación de la carrera docente, pilar de una mejora educativa.

#### **4.2. Del control estandarizado de resultados, a la mejora de procesos y de resultados educativos en forma individualizada**

La rendición de cuentas como fin de la evaluación (de los estudiantes, y por acumulación, de los sistemas) perpetúa su tradicional y largamente cuestionada función de control educativo y social, por ser instrumento de reproducción de desigualdades. Una alternativa implicaría articular esa función, a la que los sistemas educativos difícilmente renuncien por su carácter credencialista, con la de mejora del objeto a evaluar, no limitándose a la medición de resultados parciales o finales como indicadores per se, sino integrando medición y comprensión de los procesos y resultados individuales, grupales e institucionales, para su acompañamiento y mejora. En ese sentido hemos visto en 2.2. como las propias pruebas asistidas por computadora en su tercera y sobre todo en su cuarta generación ya han virado, en algunos países desarrollados, de la estandarización que homogeneiza a la fuerza, a la individualización de una evaluación que prioriza la función de acompañamiento de los procesos de aprendizaje y la autonomía de los estudiantes para aprender.

#### **4.3. La posible articulación entre la regulación de los aprendizajes y el monitoreo del sistema**

La evaluación y el monitoreo del sistema (que en nuestro país la Ley N° 18.437 asigna como cometidos al INEED<sup>16</sup>) es necesaria, y tiene como fin dar insumos para el diseño de políticas educativas.

Dada la complejidad de la educación y las insuficiencias de las pruebas estandarizadas, esa evaluación y ese monitoreo no pueden quedar reducidos a la medición de los desempeños estudiantiles por medio de ellas en forma unilateral. En ese sentido, Pedro Ravela, refiriéndose a lo que el INEED se ha propuesto, expresaba hace poco al semanario *Brecha* (30/01/15)<sup>17</sup>: «Creo que esa visión de que para saber cómo está la educación hay que ver los resultados de las pruebas de aprendizaje, que es lo que se ha hecho en la región en las últimas décadas, tiene un alcance

*limitado. (...) Para hacer política educativa necesítas mirar el conjunto de la problemática».*

Parece de orden no considerar unilateralmente ni el monitoreo del sistema solo por pruebas externas al aula (cuantitativas), cuyo criterio rector de producción es la “objetividad”, ni solo las evaluaciones de los docentes diseñadas individualmente para sus clases, que suelen ser de corte más cualitativo y son además comprensivas de una multiplicidad de dimensiones y elementos que tienen normalmente como preocupación “lo justo”. Los docentes no dejan, por lo general, de sopesar el futuro escolar del alumno y aspectos como la participación activa y el trabajo sistemático del alumno, que desde el modelo cuantitativo son cuestionados como sesgos subjetivos. Sin embargo se argumenta que como se trata de aspectos que inciden en la construcción del conocimiento por parte de los alumnos, el no tenerlos en cuenta constituye, por la negativa, un sesgo de las propias pruebas estandarizadas.

Entendemos que una política de evaluación integral debería intentar mejorar tanto las prácticas evaluatorias de los maestros como las pruebas externas, para lo cual el sistema debería orientarse a implementar las nuevas generaciones de estas últimas, articulando a su vez las lógicas diferentes de estas dos perspectivas (micro/macro) de la evaluación, contribuyendo además a elaborar reflexivamente las tensiones que surgen entre ellas<sup>18</sup> y a producir nuevo conocimiento profesional de los docentes, de las instituciones y del propio sistema.

Una formación docente universitaria, inicial y continua tendría un papel capital en ese proceso. Se requiere una formación reflexiva, crítica y teóricamente sólida sobre los dos modelos evaluatorios y sus supuestos, sus potencialidades, sobre la eventual pertinencia de su compatibilización según los objetos de conocimiento/competencias valiosos que se defina trabajar y evaluar, así como formación para el diseño de metodologías evaluatorias que confluyan en la selección y/o construcción de instrumentos de evaluación/certificación epistemológicamente consistentes e integrados al proceso didáctico.

Como ningún instrumento unilateral de evaluación da cuenta de resultados de aprendizaje y mucho menos de procesos, es necesario triangular los datos con fuentes internas (miradas de colegas, evaluaciones de pares, autoevaluaciones de los alumnos, entre otros) y externas; las pruebas estandarizadas nacionales, el intercambio con instituciones de la zona, etcétera. Se requiere, además, una formación que privilegie la investigación y el análisis de las prácticas cotidianas como parte constitutiva de la labor docente individual y colectiva.

<sup>16</sup> Instituto Nacional de Evaluación Educativa.

<sup>17</sup> En entrevista “El necio, el dedo y la luna” por Carolina Porley.

<sup>18</sup> A este respecto y solo a título de ejemplo a nivel micro, algunas opiniones de docentes sobre el instrumento pruebas en línea en la encuesta aplicada en 2012 por ANEP, que interpelan las distintas formas de descontextualización de aquellas (en tiempo, espacio, realidad curricular y pedagógico-didáctica del aula, diversidad de los alumnos), ya que, con fundamento explícito o no, la perciben como negativa. (cf. Luaces, 2014:155 [Tabla 15, “Instrumentos”]).

Es notorio, sin embargo, que no alcanza con una sólida formación, sino que se requieren diversificaciones y transformaciones en la carrera docente así como, por supuesto, remuneraciones adecuadas a los nuevos roles. Asimismo, un compromiso claro de las autoridades es necesario para valorar, posibilitar en los hechos, impulsar y apoyar el trabajo entre docentes y entre escuelas.

De acuerdo a lo precedente, pensamos que una política evaluatoria de fondo debería asignar un lugar preponderante al nivel meso (institucional), como lugar de articulación de las evaluaciones macro y micro. En ese sentido, el trabajo evaluatorio en colectivo institucional de docentes sería importante para la selección de aquellos saberes y competencias curriculares no instrumentales que corresponde priorizar y deben ser objeto de evaluación de procesos y de resultados, así como para mejorar los juicios profesionales de los maestros con relación a las evaluaciones y certificaciones referidas a sus alumnos, y para analizar e interpretar los resultados en las pruebas estandarizadas a la luz de los procesos personales realizados. El diálogo entre procesos y resultados llevaría, además, a interpretar y comprender los errores en las pruebas en forma individualizada, asumiendo su multicausalidad y no solo a partir de las posibles causas en abstracto que el dispositivo provee.

De acuerdo a Allal (2009:44-45), algunas de las tareas que podrían ser explícitamente confiadas por el Sistema a la escuela, o a un agrupamiento de escuelas trabajando en colectivo serían, por ejemplo: la elaboración de pruebas abiertas comunes para un mismo grado y el diseño del formato de los instrumentos de evaluación de procesos (por ejemplo, los portafolios empleados en una escuela); instancias de formación organizadas por la escuela sobre el rol del juicio profesional en la evaluación, tales como reuniones de seguimiento de las decisiones a tomar con relación a alumnos “en el límite” o por debajo de los criterios de suficiencia; establecer formas de concertación entre las escuelas y los liceos de la zona con relación a criterios de evaluación derivados de pautas teórico-metodológicas negociadas y concertadas, entre otras tareas que la institución considerara convenientes.

Se trata de generar conocimiento institucional que podría habilitar la realización de proyectos transformadores en la evaluación lo que, por su comprobado factor de arrastre, redundaría muy posiblemente en mejores enseñanzas personalizadas.

En resumen, como las prácticas educativas no constituyen una isla, sino que están situadas histórica y socialmente, hemos intentado abordar las pruebas estandarizadas desde un punto de vista crítico, con base en sus supuestos teóricos y ético-políticos, sus condicionamientos económicos, hoy globalizados, y algunos de sus potenciales riesgos y efectos negativos ya constatables en otros países cuando son certificativas y/o el sistema ejerce mucha presión para su aplicación, con fines de rendición de cuentas.

En una época en la que lo instrumental, desde las “app” hasta los “drones”, opaca lo sustantivo con sus indiscutibles muestras de eficiencia, nos ha parecido que podía ser de interés compartir con los colegas docentes algunas reflexiones propias y ajenas, que intentan problematizar la lógica de estos instrumentos y eventualmente resignificarlos, para integrarlos en un proceso evaluatorio abarcativo de la complejidad de los procesos y resultados de aprendizaje.

Los instrumentos, además, no son independientes del concepto de evaluación que sustentemos. Dependen tanto a nivel sistémico, institucional, como del docente, de lo que se responda a preguntas que conforman su campo: por qué y para qué evaluar; cómo se conceptualizan y priorizan los objetos a evaluar y cuáles son sus lógicas; qué evaluación/es son pertinentes en consecuencia, y cuáles son los instrumentos más adecuados para desarrollarlas.

Es en la reflexión individual y colectiva, y en el eventual diseño participativo y consensuado de una política evaluatoria, que puede cobrar sentido una evaluación orientada al logro de los fines educativos nacionales.





## Bibliografía citada

- ALLAL, Linda (2009): "Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire" (Cap. 1) en L. Mottier Lopez; M. Crahay (dirs.) (2009): *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*, pp. 29-45. Bruselas: De Boeck. Colección Pédagogies en développement.
- ANDERSON, Gary L. (2012): "Cambios institucionales, nuevos paradigmas e impacto en el trabajo docente". Ponencia en Seminario Internacional 2012. *Políticas docentes. Formación, regulaciones y desarrollo profesional*. Buenos Aires: UNESCO / IPE. Video en línea: <http://www.seminario.iipe.unesco.org/course/view.php?id=4&topic=3#top>
- BENTANCUR, Nicolás (1997-1998): "El Estado evaluador como nueva forma de relacionamiento Estado-Universidades" en *Revista Uruguaya de Ciencia Política*, N° 10. En línea: <http://www.fcs.edu.uy/archivos/20071110001818.pdf>
- BURTON, Réginald; MARTIN, Romain (2006): "L'évaluation assistée par ordinateur" en G. Figari y L. Mottier Lopez (eds.): *Recherche sur l'évaluation en éducation*, pp. 155-163. Paris: L'Harmattan. Collection Évaluer.
- FIGARI, Gérard; MOTTIER LOPEZ, Lucie (eds.) (2006): *Recherche sur l'évaluation en éducation*. Paris: L'Harmattan. Collection Évaluer.
- GIROUX, Henry (2012): *La educación y la crisis del valor de lo público. Desafiando la agresión a los docentes, los estudiantes y la educación pública*. Montevideo: Extensión Universidad de la República / Criatura Editora.
- GRUPO DE REFLEXIÓN SOBRE EDUCACIÓN (GRE) (2014): "Decimos SÍ y decimos NO. Pronunciamientos del GRE sobre la Educación". Documento N° 7. En línea: <https://vidayeducacion.files.wordpress.com/2014/12/gre-doc-nc2b0-7-nov-2014.pdf>
- HARRIS, Phillip; SMITH, Bruce M.; HARRIS Joan (2011): *The myths of standardized tests. Why they don't tell you what you think they do*. Lanham: Rowman & Littlefield Publishers.
- HURSH, David (2008): *High-stakes testing and the decline of teaching and learning. The real crisis in education*. Plymouth: Rowman & Littlefield Publishers.
- LAFONTAINE, Dominique; SOUSSI, Anne; NIDEGGER, Christian (2009): "Évaluations internationales et/ou épreuves nationales: tensions et changement de pratiques" (Cap. 3) en L. Mottier Lopez; M. Crahay (dirs.): *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*, pp. 61-80. Bruselas: De Boeck. Colección Pédagogies en développement.
- LUACES, Oscar (2014): *Evaluaciones en línea. Una innovación educativa en el marco de una computadora por alumno*. Montevideo: ANEP/Plan Ceibal. En línea: [http://www.anep.edu.uy/sea/wp-content/uploads/2014/12/Evaluaciones-en-l%C3%ADnea\\_-para-web.pdf](http://www.anep.edu.uy/sea/wp-content/uploads/2014/12/Evaluaciones-en-l%C3%ADnea_-para-web.pdf)
- MORIN, Edgar (2007): *La cabeza bien puesta. Repensar la reforma. Reformar el pensamiento. Bases para una reforma educativa*. Buenos Aires: Ed. Nueva Visión.
- MOTTIER LOPEZ, Lucie; CRAHAY, Marcel (dirs.) (2009): *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruselas: De Boeck. Colección Pédagogies en développement.
- NICHOLS, Sharon L.; BERLINER, David C. (2010): *Collateral damage. How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- PEDRÓ, Francesc (2012): "Deconstruyendo los puentes de PISA: del análisis de resultados a la prescripción política" en *Revista española de educación comparada*, N° 19, pp. 139-172. En línea: [http://www.uned.es/reec/pdfs/19-2012/06\\_pedro.pdf](http://www.uned.es/reec/pdfs/19-2012/06_pedro.pdf)
- POGGI, Margarita (2008): "De la rendición de cuentas a las responsabilidades" en *El Monitor de la Educación*. Buenos Aires: IPE/UNESCO. En línea: [http://www.buenosaires.iipe.unesco.org/sites/default/files/rendicion\\_de\\_cuentas\\_a\\_responab.pdf](http://www.buenosaires.iipe.unesco.org/sites/default/files/rendicion_de_cuentas_a_responab.pdf)
- POPHAM, W. James (1999): "¿Por qué las pruebas estandarizadas no miden la calidad educativa?" en *PREAL – Grupo de Análisis para el Desarrollo (GRADE)*. En línea: [http://www.oei.es/evaluacioneducativa/pruebas\\_estandarizadas\\_no\\_miden\\_calidad\\_educativa\\_popham.pdf](http://www.oei.es/evaluacioneducativa/pruebas_estandarizadas_no_miden_calidad_educativa_popham.pdf)
- SALES, María Teresa; RODRÍGUEZ, Luisa; SARNI, Mariana (2014): *¿Es la evaluación lo que parece? Entre el fondo y las formas*. Montevideo: Ed. Trecho.
- SCRIVEN, Michael (1967): "The Methodology of Evaluation" en R. W. Tyler; R. M. Gagné; M. Scriven: *Perspectives on curriculum evaluation*, 1. Chicago: Rand McNally.